Brain transcriptional regulatory architecture and schizophrenia etiology converge between East Asian and European ancestral populations

Sihan Liu[1], Yu Chen[1], Feiran Wang[1], Yi Jiang[1], Fangyuan Duan[1], Yan Xia[1,2], Zhilin Ning[3], Miao Li[1], Wenying Qiu[4], Chao Ma[4], Xiao-Xin Yan[5], Aimin Bao[6], Jiapei Dai[7], Richard F. Kopp[2], Liz Kuney[2], Jufang Huang[5], Shuhua Xu[3], Beisha Tang[8], Chunyu Liu[1,2], Chao Chen[1,8,9,10]★

[1] Center for Medical Genetics & Hunan Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, Changsha, Hunan 410078, China

[2] Department of Psychiatry, SUNY Upstate Medical University, Syracuse, NY 13210, USA

[3] Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China

[4] Institute of Basic Medical Sciences, Neuroscience Center, National Human Brain Bank for Development and Function, Chinese Academy of Medical Sciences; Department of Human Anatomy, Histology and Embryology, School of Basic Medicine, Peking Union Medical College, Beijing 100005, China

[5] Department of Human Anatomy and Neurobiology, Xiangya School of Medicine, Central South University, Changsha 410000, China

[6] Department of Neurobiology, Zhejiang University School of Medicine, Hangzhou 310058, China

[7] Wuhan Institute for Neuroscience and Engineering, South-Central University for Nationalities, Wuhan 430074, China

[8] National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha 410000, China

[9] Hunan Key Laboratory of Animal Models for Human Diseases, Central South University, Changsha 410000, China

[10] Hunan Key Laboratory of Molecular Precisional Medicine, Central South University, Changsha 410078, China

These authors contributed equally: Sihan Liu, Yu Chen, Feiran Wang.

★Corresponding author: email chenchao@sklmg.edu.cn

Abstract

Understanding the genetic architecture of gene expression and splicing in human brain is critical to unlocking the mechanisms of complex neuropsychiatric disorders like schizophrenia (SCZ). Large-scale brain transcriptomic studies are based primarily on populations of European (EUR) ancestry. The uniformity of mono-racial resources may limit important insights into the disease etiology. Here, we characterized brain transcriptional regulatory architecture of East Asians (EAS; n=151), identifying 3,278 expression quantitative trait loci (eQTL) and 4,726 spliceQTL (sQTL). Comparing these to PsychENCODE/BrainGVEX confirmed our hypothesis that the transcriptional regulatory architecture in EAS and EUR brains align. Furthermore, distinctive allelic frequency and linkage disequilibrium impede QTL translation and gene-expression prediction accuracy. Integration of eQTL/sQTL with genome-wide association studies reveals common and novel SCZ risk genes. Pathway-based analyses showing shared SCZ biology point to synaptic and GTPase dysfunction as a prospective pathogenesis. This study elucidates the transcriptional landscape of the EAS brain and emphasizes an essential convergence between EAS and EUR populations.

Main

Population genetics examines differences within and between populations and how such genetic differences contribute to health and disease. A global understanding of the influence of genetic variance on complex diseases would advance insight into the biological mechanisms of disease risk[1]. During the past decade, genome-wide association studies (GWAS) have identified thousands of risk variants for psychiatric disorders across diverse populations[2,3]. Nonetheless, most of samples in psychiatric disorders GWAS originate from those of European (EUR) descent[4]. Due to ancestral differences evident in allele frequencies (AF), linkage disequilibrium (LD) patterns, and other factors, GWAS findings often fail to translate to other populations[5,6]. For example, Martin et al. examined that genetic risk prediction accuracy will decrease within heterogeneous populations which the original GWAS sample and target of prediction are divergent[5].

Interpreting GWAS “hits” with expression quantitative trait loci (eQTL) and splicing quantitative trait loci (sQTL), significantly enriched for trait-associated SNPs offers a feasible alternative for advancing our understanding of the molecular mechanisms underlying complex traits[7,8]. In the past decade, eQTL and sQTL have become familiar and effective tools enabling GWAS to explain single nucleotide polymorphism (SNP) heritability and spotlighting potential disease risk genes[9-14]. Various methods have been proposed to interpret GWAS using eQTL/sQTL signals to establish gene-expression prediction models. Examples such as PrediXcan[15] and TWAS[16] correlate imputed gene expression to a phenotype under investigation. However with eQTL/sQTL, the problem of population disparity becomes even more extreme, as most resources focus largely on the EUR ancestry alone[17-20]. The capacity of existing prediction models to isolate causal genes common across populations appears to be constrained by the Eurocentricity of the models themselves. In this way, the limited availability of non-EUR GWAS impedes our ability to fully understand the genetic basis

of diseases and to translate basic research into clinical medicine.

Recent studies have discovered significant inefficiency in predictive performance between heterogenous populations[14,21-23]. One plausible explanation for this is the differences in LD patterns and AF distribution. These disparities also hinder the ability of QTL to replicate in diverse populations. For example, comparing the regulatory architecture of gene expression in lymphoblastoid cell lines, Stranger *et al.* found that QTL differentiation among populations was likely due to AF differences reducing the statistical power of association testing[24]. Additionally, Lauren *et al.* showed that the AF differences between populations led to the accurate prediction of some genes and poor prediction in others[22]. Therefore, developing new transcriptome regulatory profiles and prediction models specific to ancestral populations is critical for accurately predicting gene expression and identifying disease risk genes.

It should be noted that transcriptomic studies conducted for other tissues (e.g., blood), cannot adequately represent the transcriptome of neuropsychiatric disorders that are most closely associated with the brain[25]. Gene expression is tissue-specific. Moreover, many studies have discovered that QTLs within specific pathogenic tissues are significantly enriched for relevant trait associations[26-30]. For example, in the frontal cortex, a region widely accepted as critical for schizophrenia (SCZ), QTLs detected are significantly enriched with greater SCZ GWAS signals than QTLs detected from other tissues[26-28]. Such findings signal the need to develop regulatory profiling of the human brain to uncover the biological mechanisms of SCZ. Several studies have generated large-scale postmortem brain data[31-33]. For instance, Wang *et al.* developed a comprehensive resource for functional genomics of 1,866 adult brains using PsychENCODE data that highlights key genes and pathways associated with SCZ, including immunological, synaptic, and metabolic pathways[33]. To our knowledge, no systematic investigation into whether the genetic control of gene expression and splicing in brain is similar or varies between populations exists. Furthermore,

102 determining whether differences represent etiologic heterogeneity in SCZ across
103 populations also begs investigation.

104 Here, we developed a novel brain transcriptome dataset comprised of 151 EAS
105 individuals. We hypothesize that the brain's regulatory architecture of gene expression
106 and the etiology of SCZ converge between EAS and EUR populations. We also posit
107 that AF and LD patterns distinct to ancestral populations are at least in part responsible
108 for QTL heterogeneity across populations. To test these hypotheses, we applied eQTL,
109 sQTL and co-expression analyses, comparing the results with existing
110 PsychENCODE/BrainGVEX data (specifically the EUR subpopulation) to evaluate the
111 similarities and differences in the transcriptional regulatory architecture of the two
112 populations. By integrating eQTL and sQTL results with SCZ GWAS summary data,
113 we quantified the enrichment of eQTL/sQTL associated with GWAS signals. Further,
114 we identified common and novel risk genes as well as disease-related pathways. From
115 these data we assembled a new genome-wide human brain regulatory map, which
116 affords considerable insight into the biological progression of SCZ in East Asians.

Results

To identify EAS-specific regulatory variants shaping brain gene expression and alternative splicing, we performed high-density genotyping and high-throughput RNA-sequencing in 151 EAS prefrontal cortices (Fig. 1). After quality control and preprocessing (Methods and Extended Data Fig. 1), we gathered 18,939 brain-expressed genes and 6.4 million autosomal SNPs. PCA (principal component analysis) of ancestry verified the East Asian ethnicity of all donors (Supplementary Note). eQTL and sQTL mapping and constructed co-regulatory networks enabled us to examine the brain expression regulatory architecture of each population individually, with the EUR population derived from the PsychENCODE/BrainGVEX project.

Identifying and characterizing the function of cis-acting expression QTLs and splicing QTLs revealed common enrichment patterns between populations

We identified cis-eQTLs using QTLtools[34](Fig. 1), adjusting for 20 hidden covariates identified by the probabilistic estimation of expression residuals (PEER)[35] (Methods and Supplementary Note). These hidden factors were significantly correlated with technical and biological covariates such as experimental batch, RNA Integrity Number (RIN), sex, and age of death. We identified 3,278 genes with a cis-eQTL (false discovery rate (FDR) q-value < 0.05) in EAS populations, 10,043 genes with a cis-eQTL (FDR q-value < 0.05) in EUR populations, which are referred to as eGenes (Table 1).

By identifying numerous excised intronic clusters using LeafCutter[36] (Methods and Fig. 1), we were able to discover sQTLs as well. We identified 4,726 significant sQTLs (FDR q-value < 0.05) in EAS and 18,927 significant sQTLs (FDR q-value < 0.05) in EUR populations, which were mapped to 2,054 and 5,641 genes (sGenes) respectively per population (Table 1).

To better characterize the function of the eQTLs and sQTLs, we evaluated their

143 distance distribution and enrichment in numerous functional regions. Our first finding
144 agreed with previous results conducted in EUR brains[33,37,38]: 20% of the eQTLs in both
145 populations were located within 10kb of the transcription start site (TSS) regions (Fig.
146 2a,b); the most significant ($FDR_{permutation}$ q-value <0.05) SNP per sQTL (sSNP) showed
147 clustering around the splice junction. Fifty percent of sQTLs are located within 10 kb
148 of the splice junction (Fig. 3a,b) in both EAS and EUR populations, demonstrating that
149 variants proximal to splicing junctions have a large effect. In contrast to eQTL, the
150 majority of sSNPs (60%) lie within the gene body (Fig. 3c), also consistent with
151 previous research[38].

152 We then annotated expressed SNPs (eSNPs) and sSNPs with chromatin state
153 predictions for prefrontal cortical tissue using GREGOR[39] (Methods). We found that
154 eSNPs and sSNPs were significantly enriched in the same TSSs, promoters, and
155 transcribed regulatory promoters or enhancers ($P_{Bonferroni}$ < 0.05, Fig. 2c, Fig. 3d;
156 Supplementary Table 4). We also annotated eSNPs with transcription factor binding
157 sites (TFBS) and sSNPs with experimentally determined RNA binding protein (RBP)
158 binding sites. We observed that 46 and 49 TFBS were significantly enriched with
159 eQTLs in the EAS and EUR populations separately ($P_{Bonferroni}$ < 0.05, Fig. 2d,e and
160 Supplementary Table 4). All of TFBS that were significantly enriched with eQTLs in
161 EAS population were also significantly enriched with eQTLs in EUR population.
162 Furthermore, sQTLs were significantly enriched in binding targets of 7 RBPs in the
163 EAS population, while binding targets of 71 RBPs were significant in the EUR sQTL
164 dataset ($P_{Bonferroni}$ < 0.05, Fig. 3e,f and Supplementary Table 4). Five of the seven RBPs
165 that were significantly enriched with sQTLs in the EAS population were also
166 significantly enriched with sQTLs in EUR populations.

167

168

Table 1: Identification of eQTLs and sQTLs

	#SNPs	#Genes	#Intron clusters	#cis-eQTLs/sQTLs	#Significant eGenes/sGenes
EAS(n=145)	6,045,349	18,939	146,884	604,485/994,668	3,278/2,054
EUR(n=397)	8,108,028	16,542	188,310	2,790,193/4,705,755	10,043/5,641
Across population	4,681,303	16,266	132,619	286,288/442,281	2,650/1,779

Cis-eQTLs/sQTLs are defined as FDR q-value < 0.05. Significant cis-eQTLs/sQTLs across populations defined as eQTLs/sQTLs significant in EAS and EUR populations in the same direction. Significant eGenes are genes regulated by SNPs that passed multiple testing (FDR q-value < 0.05) based on the permutation-based analysis. Significant sGenes are genes that intron clusters can map to and passed multiple testing (FDR q-value < 0.05) based on the permutation-based analysis.

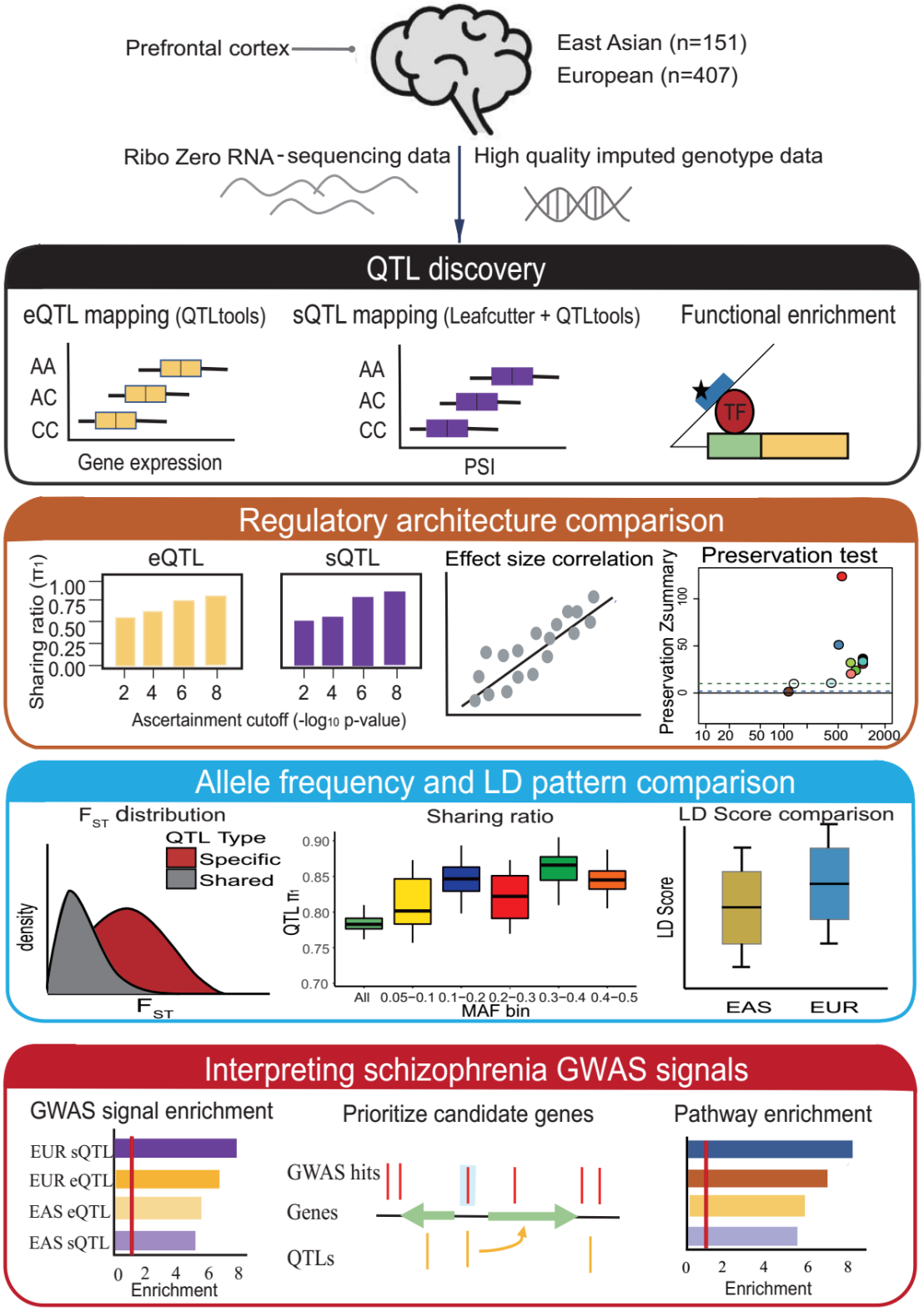


Fig. 1 | Study design. We collected genotype and RNA-seq data from East Asian (n = 151) and European populations (n = 407). After quality control and data preprocessing, eQTL and sQTL were independently calculated using standard

179 methods of covariate correction. QTLs were characterized based on functional
180 enrichment. Then, we compared the regulatory architecture including eQTL/sQTL and
181 the gene co-regulatory patterns between EAS and EUR populations. Next, we
182 calculated F_{ST} and LD scores to evaluate the contribution of AF and LD patterns
183 difference in QTL comparison. Finally, to determine whether schizophrenia biology
184 between East Asian and European populations is analogous, we integrated QTLs
185 previous identified with SCZ GWAS to identify disease risk and important biological
186 processes under genetic control. PSI: percent-spliced-in; F_{ST}: Fixation index,
187 measures the population differentiation due to genetic structure; LD: linkage
188 disequilibrium; LD score: the sum of the LD r^2 between the focal SNP and all the
189 flanking SNPs within a 1cM window with 1000G data.

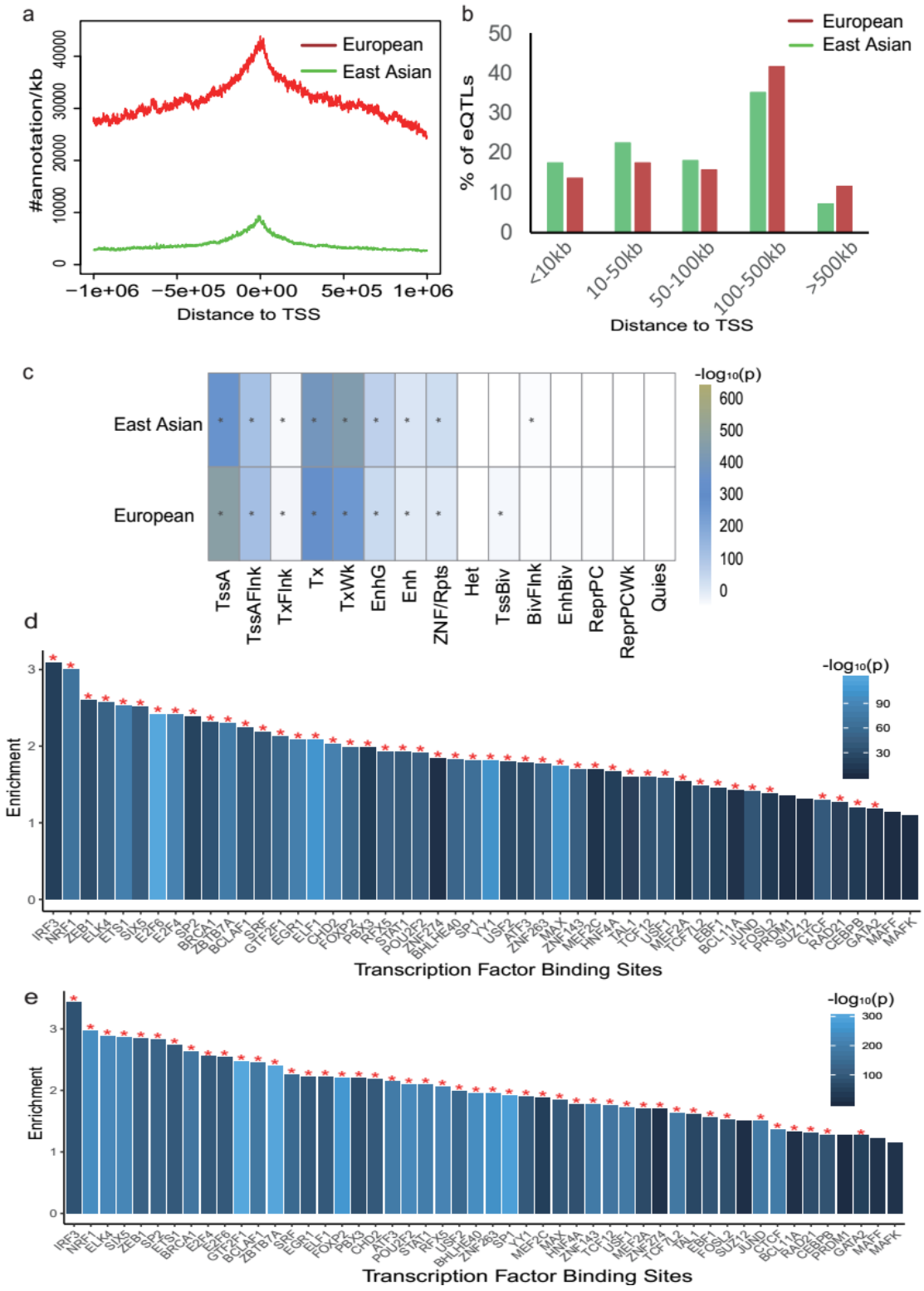


Fig. 2 | Characterization of eQTLs. a, Distance distribution of eQTLs in the East Asian (green) and European (red) populations to the TSS as defined in Gencode v19. **b,** Percentage of distance distribution of all cis-eQTLs in East Asian (green) and European (red) populations. **c,** Enrichment of eSNPs in 15 core models. eSNPs in both

195 populations most significantly enriched in the TSSs, promoters, and transcribed
196 regulatory promoters or enhancers. *$P_{Bonferroni} < 0.05$. **d,** Enrichments of eSNPs in
197 experimentally discovered transcription factor binding sites in the East Asian
198 population. **e,** Enrichments of eSNPs in experimentally discovered transcription factor
199 binding sites in the European population. *$P_{Bonferroni} < 0.05$.

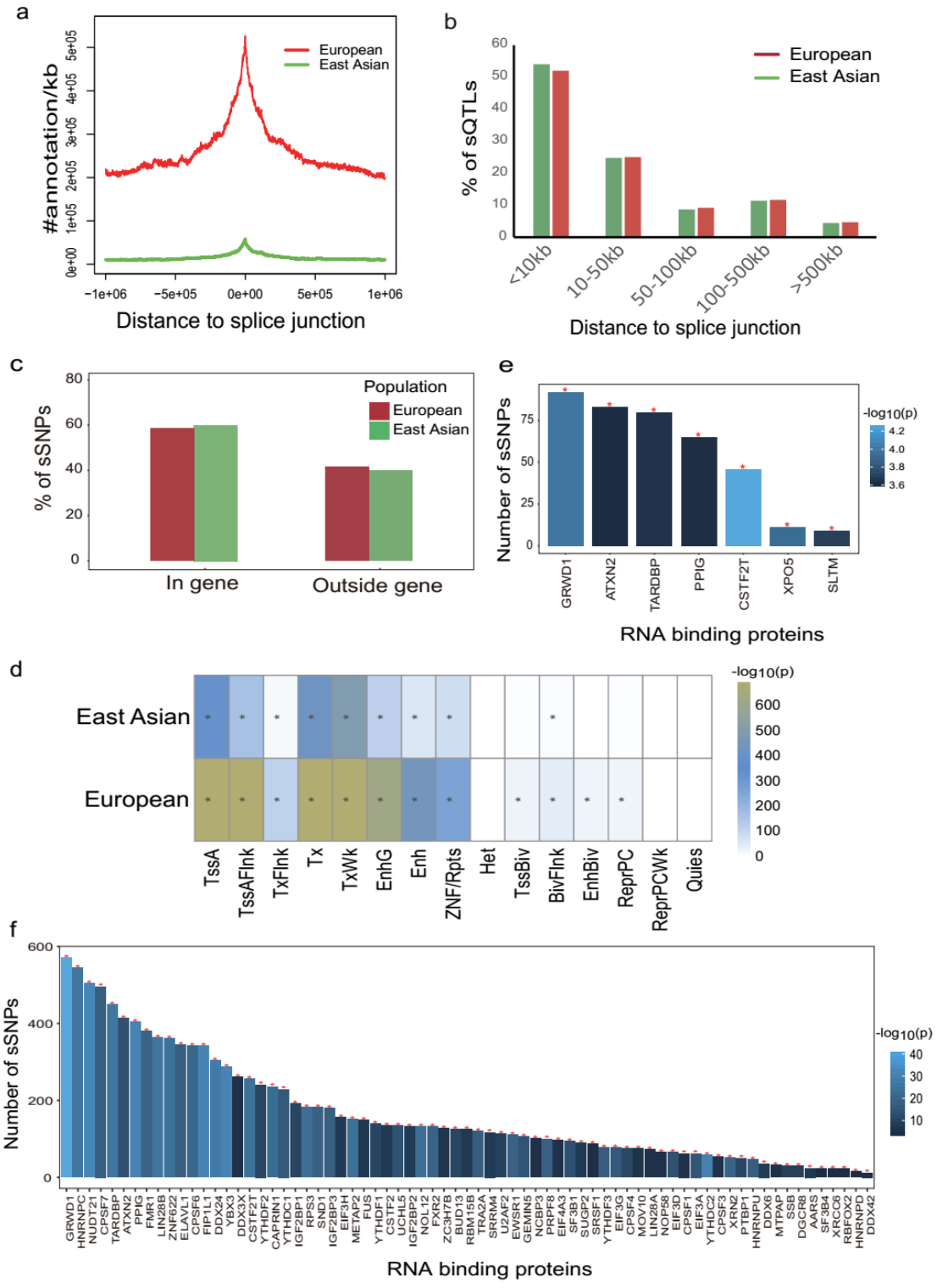


Fig. 3 | Characterization of sQTLs. a, Distance distribution of sQTLs to the splice junction. sQTLs from the East Asian (green) and European (red) populations. **b,** Percentage of distance distribution of all cis-sQTLs in East Asian (green) and European (red) populations. **c,** Fraction of sQTLs where the sSNP lies within vs outside

its sGene. **d,** Enrichment of sSNPs in 15 core models. *$P_{Bonferroni} < 0.05$. **e,** RBP enrichment among the significant sSNPs in the East Asian population. **f,** RBP enrichment among the significant sSNPs in the European population. *$P_{Bonferroni} < 0.05$.

Brain expression regulatory architectures are broadly preserved across EAS and EUR populations

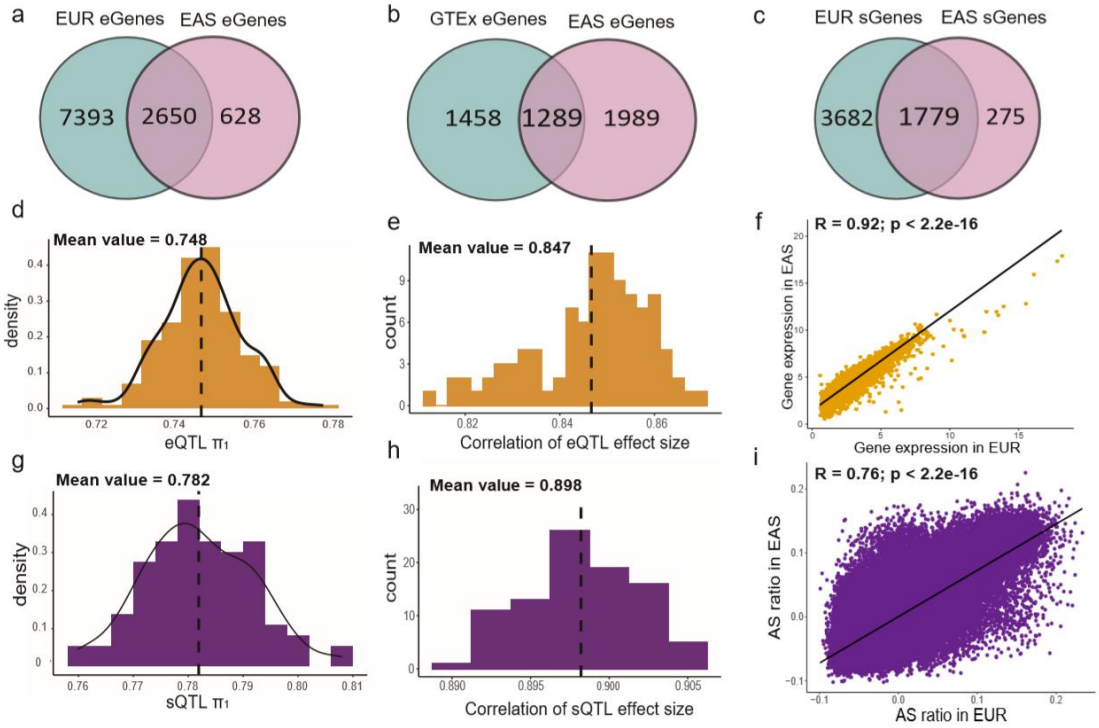
An important aim in this study is to investigate to what degree the genetic control of gene expression and splicing in the brain varies between human ancestral populations. We first compared eGenes identified in EAS and EUR populations and found that 2,650 eGenes overlapped. Most (80%) eGenes detected in the EAS population were also significant in the EUR population (Fig. 4a). Additionally, we compared the eGenes detected in the EAS population to those from GTEx adult cortices[40]. We found 1,289 overlapping eGenes, accounting for nearly 40% between both datasets (Fig. 4b). We also compared sGenes across EAS and EUR populations. Results showed 1,779 overlapped sGenes (Fig. 4c), 87% of which were shared significantly across populations.

We next used Storey's π_1 statistic to assess the extent of eQTL/sQTL sharing across populations. To assess the true extent of this sharing, we performed down-sampling analysis with 100 repetitions (Methods). The fraction of eQTLs and sQTLs shared between the EAS and EUR populations was 74.8% and 78.2%, respectively (Fig. 4d,g). Moreover, we calculated the Pearson correlation of genetic effect size between shared QTLs and found that the genetic effect size between EAS and EUR populations was highly analogous (R_{eQTL}=0.847; R_{sQTL}=0.898; Fig. 4e,h).

We completed a meta-analysis, pooling results from diverse populations, to gain greater statistical power for QTL detection and to identify shared QTLs across populations. We calculated a meta p-value using METAL[41] for each eQTL/sQTL across populations; eQTLs or sQTLs at a meta FDR < 0.05 were referred to as 'meta

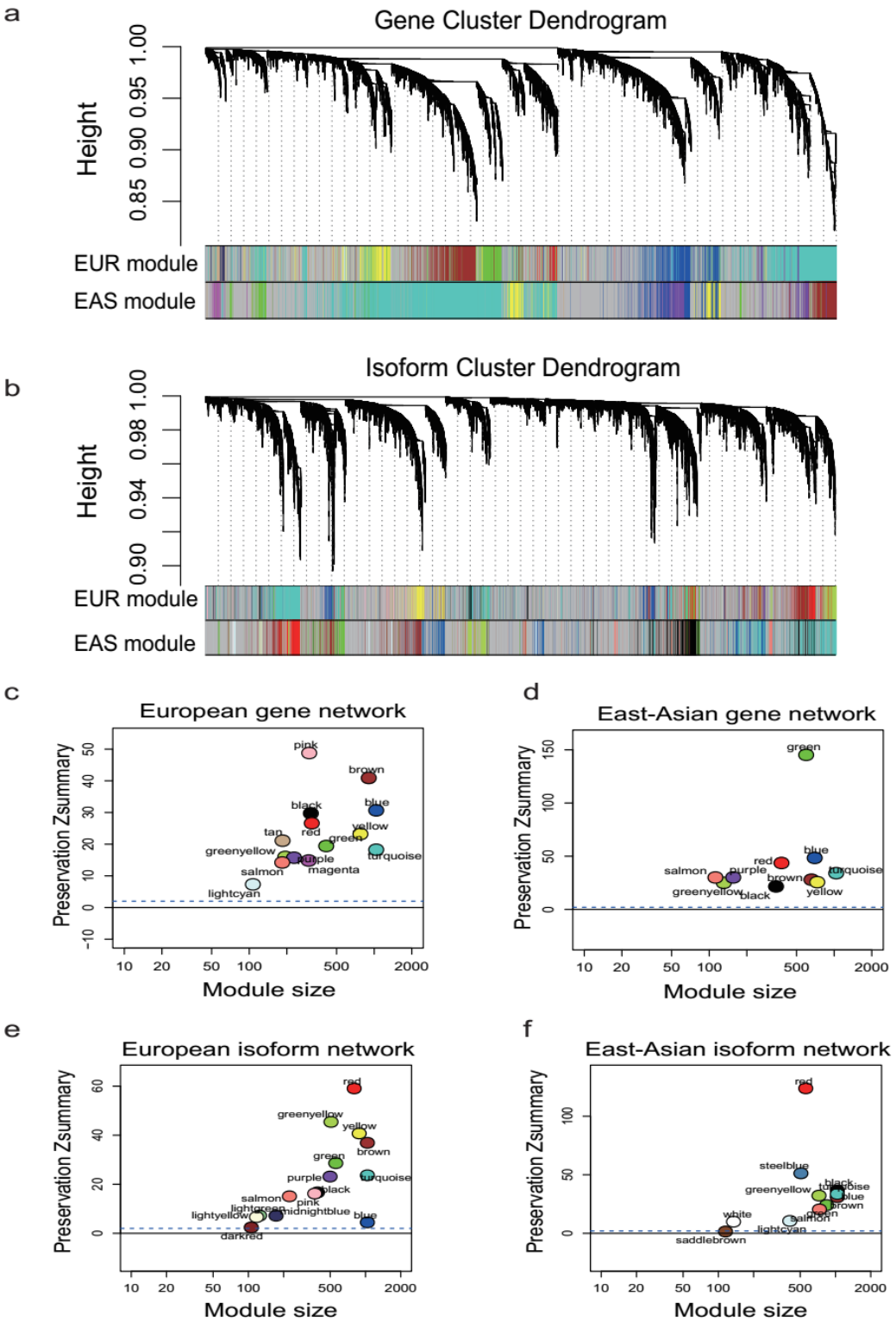
eQTLs/sQTLs'. Greater than 80% of these were significant across populations and showed concordant regulatory direction across populations (Extended Data Fig. 2 and Supplementary Table 2). Also, numerous new eQTL/sQTL signals were identified by meta-analysis.

To comprehensively compare the brain expression regulatory architecture between EAS and EUR populations, we calculated the Pearson correlation of gene expression between the two populations using shared genes. We found that gene expression was highly correlated in the two populations (Fig. 4f; R = 0.92, p-value < 2.2e-16), and a similar result was observed for alternative splicing ratio (Fig. 4i; R = 0.76, p-value < 2.2e-16). Furthermore, we applied weighted gene co-expression network analysis (WGCNA)[42] and robust WGCNA to create independent gene- and isoform-level networks. We then used preservation testing to evaluate the consensus of networks constructed by each population. Preservation Z summary score of each module was >2 in both the gene expression and isoform levels, showing that co-expression patterns are broadly preserved between EAS and EUR populations (Fig. 5c,d,e,f and Supplementary Table 5).



247

248 **Fig. 4 | Comparison of the regulatory pattern. a,** Venn Diagram for eGenes
249 discovered in European (EUR) population vs East Asian (EAS) population. **b,** Venn
250 Diagram for eGenes discovered in adult cortical tissue from GTEx vs EAS population.
251 **c,** Venn Diagram for sGenes discovered in EUR population vs EAS population. **d,**
252 Distribution of eQTL π_1 between EAS and down-sampled EUR populations. The mean
253 π_1 value was 0.748. **e,** Distribution of correlation coefficient between eQTL effect size.
254 The mean correlation coefficient value was 0.847. **f,** Pearson's correlation in expressed
255 genes between EAS and EUR populations. **g,** Distribution of sQTL π_1 between EAS
256 and down-sampled EUR populations. The mean π_1 value was 0.782. **h,** Distribution of
257 correlation coefficient between sQTL effect size. The mean correlation coefficient value
258 is 0.898. **i,** Pearson's correlation in intron clusters between EAS and EUR populations.
259 AS: alternative splicing.



260

261 **Fig. 5 | Comparison of co-expression pattern**. **a,** Network analysis dendrogram

262 based on hierarchical clustering of all genes by their topological overlap. Colored bars

263 below the dendrogram show module membership. **b,** Network analysis dendrogram

based on hierarchical clustering of all isoforms by their topological overlap. **c,** Preservation Z summary score of each gene co-expression module in the EUR population. The x-axis is the number of genes in each module and the y-axis is Z summary score, which measures the preservation between modules. When Z summary score >=2, it indicates that this module was preserved in another population. The blue dashed line is the moderately conserved threshold. Each point represents a module constructed in population, labeled by color. **d,** Preservation Z summary score of each gene co-expression module in the EAS population. **e,** Preservation Z summary score of each isoform co-expression module in the EUR population. **f,** Preservation Z summary score of each isoform co-expression module in the EAS population.

Differences in AF and LD across populations decrease the QTL reproducibility and the power of gene expression prediction

Although the sharing ratio reached almost 80%, 20% of the QTLs are significant only in one population. We hypothesized that part of QTL differentiation is due to population divergence in AF and LD pattern across populations. To address this hypothesis, we defined EUR robust QTLs as significant QTLs detected at least 50 times in down-sampling analysis, ancestry-specific QTLs as significant in one population, and ancestry-shared QTLs as significant in both populations. We identified each of these QTL types by comparing the lists of EAS QTLs and EUR robust QTLs. To investigate the effect of AF on QTL differentiation, we estimated F_{ST} (fixation index) for each eSNP and sSNP and compared the distribution of F_{ST} between ancestry-specific and ancestry-shared QTLs (Methods). We found that ancestry-specific QTLs were significantly enriched in population-divergent SNPs (F_{ST} > 0.05; Fisher exact test: P < 2.2e-16) and ancestry-shared QTLs were significantly enriched in population-convergent SNPs (F_{ST} < 0.05; Fisher exact test: P < 2.2e-16; Fig. 6a). To further verify our hypothesis, we separated eSNPs/sSNPs into different minor allele frequency (MAF) bins, and calculated QTL π_1 in each bin. We found that with similar AF, QTL sharing

ratios were higher than cross SNPs with different AF, suggesting that the SNPs with less population divergence were more likely to be eQTLs/sQTLs shared by the two populations (Fig. 6b,c).

We then tested whether ancestry-specific QTL loci have unique LD patterns (Methods). Results showed LD patterns for ancestry-specific QTLs varied significantly between EAS and EUR populations (Wilcoxon tests, $P < 0.05$; Fig. 6d). Further, correlation coefficients between F_{ST} and LD score were less than 0.1 ($P < 0.01$), suggesting that the cross-population differences in LD patterns affect QTL differentiation independently when compared with F_{ST}.

Associations between SNPs and genes enable the development of predictive models that can “impute” gene expression when phenotype-related tissue types are unavailable. However, population-specific QTL signals may reduce the accuracy of gene expression prediction across populations. We hypothesize that prediction performance will be lower when a model trained on one population is used to predict gene expression in another population. To investigate, we compared gene-expression predictive performance within and across EAS and EUR populations. We used matched SNPs and genes in both populations (n=145) to build the models (n = 100) using PrediXcan[15] (Methods). The Pearson correlation between predictive performance of genes in the EAS and EUR populations was 0.60 (Fig. 6e). We also found that single-population-trained models had significantly decreased performance when predicting gene expression in another population (Wilcoxon tests, EAS Model: $P = 4.577 \times 10^{-12}$; EUR Model: $P < 2.2 \times 10^{-16}$; Fig. 6f; Supplementary Table 3)

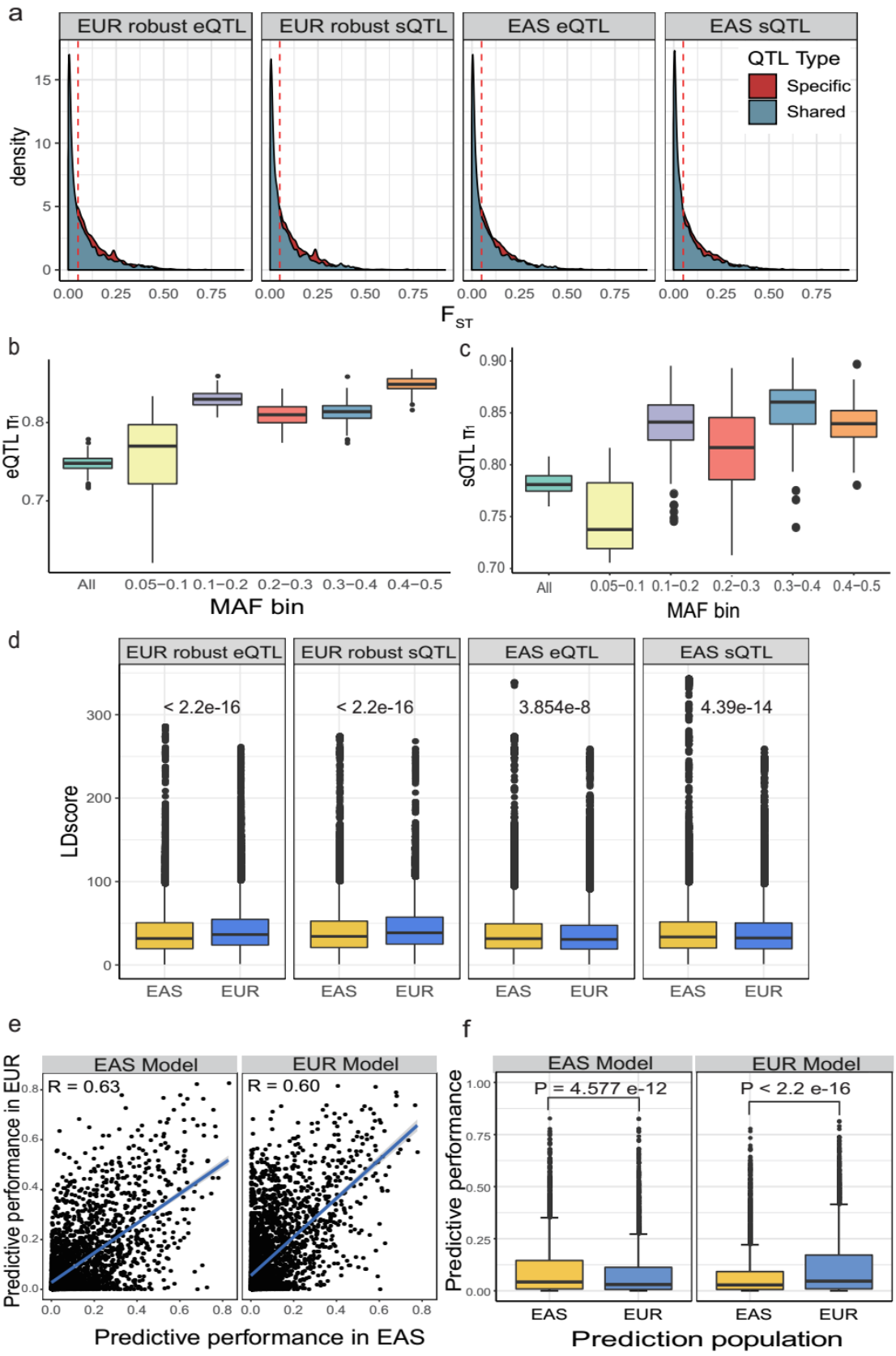


Fig. 6 | AF and LD differences contribute to QTL differences between populations.

a, Comparison of F_{ST} between ancestry-shared and ancestry-specific QTLs. Robust

316 QTL: detected as significant QTLs at least 50 times in down-sampling analysis. **b,**
317 eQTL π_1 in different MAF bins. **c,** sQTL π_1 in different MAF bins. **d,** Linkage
318 disequilibrium score distribution comparison for ancestry-specific QTLs between EAS
319 and EUR populations. **e,** Comparison of predictive performance for each gene (R^2)
320 between EAS and EUR populations in different prediction models (EAS and EUR
321 model). The identity line is shown in blue. **f,** Comparison of predictive performance
322 between genes in different prediction models.

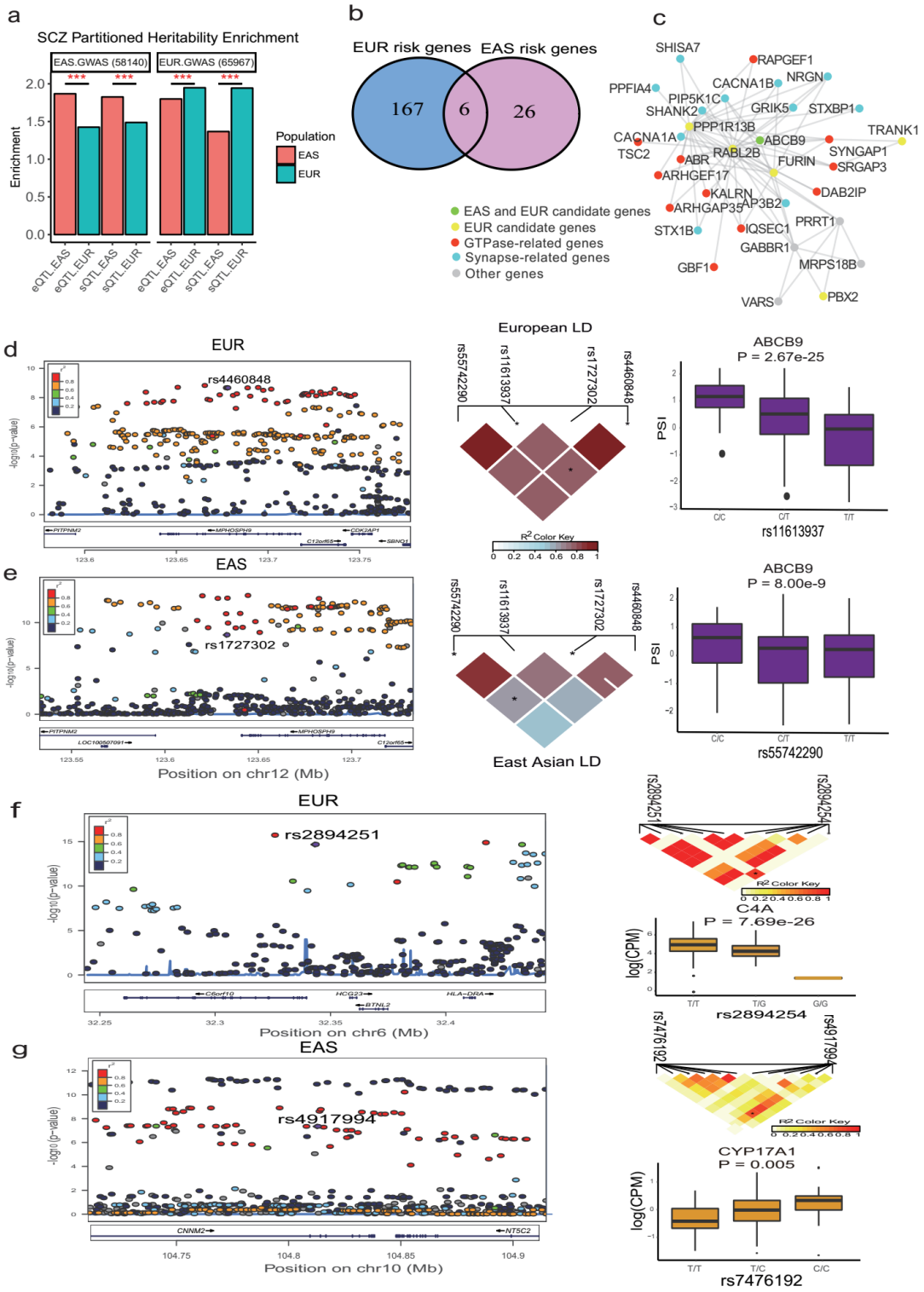
323 **Synapse- and GTPase-related pathway implicated in SCZ risk across populations**

324 eQTLs and sQTLs detected in the human brain can help to decipher the unlock
325 biological mechanisms of SCZ. To examine whether QTL results from the phenotype-
326 linked population explains more disease signals and SNP heritability than those from
327 disparate populations, we first collected SCZ GWAS summary statistics for both
328 ancestral populations from the Psychiatric Genomics Consortium (PGC)[43,44]. We then
329 compared GWAS signal enrichment using partitioned LD-score regression (LDSR)[45]
330 (Methods). For the EAS GWAS summary data, eQTLs/sQTLs detected in the EAS
331 population showed greater significant enrichment in GWAS signals than those from
332 the EUR population (Welch Modified Two-Sample t-Test $P < 0.001$) and vice versa
333 (Welch Modified Two-Sample t-Test $P < 0.001$; Fig. 7a and Supplementary Table 6).
334 We corrected for possible sample size variance bias by calculating the enrichment of
335 robust EUR QTLs, and the results agreed with previous reports (Extended Data Fig.
336 3a,b).

337 It was next necessary to evaluate whether observed differences represented true
338 etiologic heterogeneity of SCZ across populations. To achieve this, we used regulatory
339 trait concordance (RTC)[46] and summary data-based Mendelian randomization (SMR)[9]
340 to prioritize SCZ candidate risk genes (Methods). We prioritized 199 SCZ candidate
341 risk genes, including 173 genes in the EUR population and 32 in the EAS population
342 (Supplementary Table 7; Supplementary Table 8 and Fig. 7b). Six of the 199 were

identified in both populations (CNNM2, C12orf65, MPHOSPH9, MARCKSL1P1, C2orf47, and ABCB9). Other genes were identified within a single population. For example, C4A was identified as a risk gene in the EUR population (Fig. 7f) by integrating the eQTL signals, while CYP17A1 was identified as a risk gene in the EAS population by integrating the eQTL signals (Fig. 7g). Comparing published SCZ risk genes with the 199 candidate genes we identified, 77 of the EUR candidate genes and 10 of the EAS candidate genes aligned (Methods and Supplementary Table 9).

Along with peripheral genes, these candidate genes form a network that fulfills specific functional roles. To better characterize the biological function of these candidate genes, we analyzed the enrichment of candidate genes having previously constructed networks (Methods). We tested whether modules were significantly enriched in candidate genes for both EAS and EUR populations, but none were. We also tested whether any consensus modules (preserved in both populations) were significantly enriched with candidate genes present in both populations. One consensus module was significantly enriched in candidate genes from the combined populations (Fig. 7c; p-value = 0.01). Enrichment analysis showed function related to synapse and GTPase pathways, including regulation of chemical synaptic transmission, neuron projection development, synapse structure or activity, small GTPase mediated signal transduction, and GTPase binding (Extended Fig. 3d and Supplementary Table 10).



362

363 **Fig. 7 | Explanation of SCZ GWAS signals and prioritization of candidate genes.**

364 **a,** GWAS enrichment results from LDSR. ***: Welch Modified Two-Sample t-Test P <

365 0.001. **b,** Venn plot for SCZ risk genes in EUR and EAS populations with combined

366 RTC and SMR results. **c,** The sub-network of yellow module. **d-e,** Examples of shared

SCZ risk genes. **f,** Example for EUR-specific risk gene, C4A. **g,** Example for EAS-specific risk gene, CYP17A1. LDSR: LD score regression.

Discussion

We developed a novel brain transcriptome data set and compiled the first genome-wide brain regulatory map of the prefrontal cortex from a solely EAS population. We identified 3,278 eQTL and 4,726 sQTL signals that reached a genome-wide level of significance. Detected eSNPs and sSNPs corresponded to previous reports[38,40,47,48] with significant enrichment in active functional regions such as promoters and enhancers. Comparing the EAS data with the PsychENCODE/BrainGVEX-derived EUR data, we found most regulatory elements common to both populations. Moreover, by integrating QTL signals with summary statistics from SCZ GWAS, we observed synapse- and GTPase-related pathways involved in the development of SCZ in both populations.

This study demonstrated convergent transcriptional regulatory architectures between EAS and EUR populations through multiple lines of evidence. Meta-analysis revealed approximately 80% of QTLs were shared between populations. Moreover, several relational analyses suggest a high degree of congruence between EAS and EUR populations, including the π_1 statistic (eQTL=0.748; sQTL=0.782), correlations between populations for genetic effect size (eQTL=0.847; sQTL=0.898) and correlations for gene expression and co-regulatory networks. Our study of post-mortem brain tissue concurs with studies based on whole blood and liver tissue in which EAS and EUR cis-eQTL replication rates equaled 60%[49] and 40%[50] respectively. Our study parallels previous comparisons of the genetic control of gene expression[24,50,51], methylation[52], and chromatin accessibility[53] generated from lymphoblastoid cell lines in diverse populations from worldwide reference panels[54-56], showing that regulatory patterns are shared across populations.

Our proposal that some QTLs have population-specific effects is not unique [24,53]. Seeking to find the biological mechanisms underlying these divergent effects, we compared F_{ST} and LD score distribution for ancestry-specific QTLs across populations. Here, we found significant differences. These results indicate that a degree of QTL differentiation signals divergence in AF and LD. Evidence suggests that such genetic differentiation in ancestral populations is due primarily to natural selection[57,58]. Besides, contemporary populations descend from dramatically smaller migratory populations (bottleneck effect), hence, population-specific QTLs could arise from bottleneck effect and environmental factors including climate, diet, and pathogenic microorganisms.

Population-specific QTLs have important implications for predictive modeling. QTL signals lay the foundation for predictive models and assist in imputing gene expression when tissues relevant to phenotypes are unavailable. Our gene prediction model that was trained in one population decreased their prediction performance when predicting gene expression for other populations at a ratio of 14% to 33% respectively. This agrees with several recent studies reporting superior accuracy of prediction models in target populations with ancestry comparable to the discovery population[14,21-23]. Therefore, population-specific predictive models are integral for transcriptome mapping of the human brain.

Population-specific regulatory regions may harbor a portion of the disease risk. This may limit QTL's utility in interpreting GWAS signals in disparate populations. We compared the enrichment of QTL signals in SCZ GWAS across populations and found more significant enrichment of eQTLs/sQTLs in the discovery population than in the disparate populations. Similar results have been reported in Type 2 diabetes[59]. These findings highlight the importance of using GWAS to interpret QTLs from the target population in accurately explaining the disease signals within that population.

Since SCZ occurs with similar prevalence and a genetic basis broadly shared across populations[43,60], we would expect distinct groups to share many risk genes.

Surprisingly, we observed only six of the 199 SCZ risk genes in common between EAS and EUR populations. Nevertheless, this finding should not be interpreted to mean that EAS and EUR populations carry entirely different risk genes. For, we found that almost 70% of the SNPs that are linked to regulating population-specific SCZ risk genes vary in AF and LD patterns across populations. For example, we identified the EUR-specific risk gene C4A, a target of extensive scrutiny in association with SCZ[61-63]. C4A localizes to the MHC class III region on chromosome 6, which is strongly associated with SCZ and which hosts a EUR-specific LD pattern. The corresponding GWAS signal rs2894251 was significant within the EUR population (P = 2.144 e-15; MAF = 0.12) but not so in the EAS population (P = 0.05156; MAF = 0.02). Although the associated eSNP rs2894254 is extremely uncommon in the EAS population (MAF < 0.001), it has an MAF of 6% in the EUR population. These results suggest that at least some differences in EAS and EUR SCZ risk genes are due to low AF and disparate LD patterns, which may account for the loss of risk genes. Risk genes were more readily detectable in both populations when they were present in similar or higher AFs with similar LD patterns. Yet, many SCZ risk genes are evident within low frequencies too, hindering consistent detection.

It is well known that complex molecular networks and cellular pathways fuel disease susceptibility and development[64,65]. Therefore, we exploited pathway enrichment analyses of identified risk genes and co-regulated genes to explore the mechanisms behind SCZ. SCZ risk genes were significantly enriched in one consensus module (module yellow) for both ancestral populations. This module was enriched for an array of established SCZ modular pathways, including synapse- and GTPase-related pathways[66-70]. Pathways related to neuron-to-neuron, postsynaptic density, and asymmetric synapses are established suspects in genetic risk of SCZ. A meta-analysis showed a significant decrease in the density of postsynaptic elements in SCZ patients compared to healthy controls[69]. GTPase-related pathways, including regulation of

small GTPase-mediated signal transductions and GTPase binding, have also been implicated in SCZ. A previous study shows that a missense polymorphism (H204R) of a Rho GTPase-activating protein is associated with schizophrenia in men[70]. Our results support the premise that synapse-related and GTPase-related pathways have an important role in the etiology of schizophrenia for both EAS and EUR populations.

Our study showed how both eQTLs and sQTLs benefit the study of underlying disease mechanisms. We discovered how heritability explained by eQTLs and sQTLs is similar in both EAS and EUR populations. Recent studies have examined the contribution of regulatory variants to SCZ, educational attainment, and autism spectrum disorder (ASD), concluding that sQTLs contribute comparably or with even greater magnitude than eQTLs[8,14,38]. Additionally, we found that although only 14% of SCZ risk genes were identified by both eQTL and sQTL signals (Extended Data Fig. 3c,d), 40% of the SCZ risk genes identified by integrating either eQTLs or sQTLs had also been reported as SCZ risk genes in previous literature[66,71-77]. This result indicates that eQTLs and sQTLs can identify distinct risk genes which facilitate our understanding of disease mechanisms.

In general, our results show the transcriptional architecture of expression regulation and the underlying SCZ biology converging between the EAS and EUR populations. Synaptic- and GTPase- related pathways are likely suspects in the pathogenesis of SCZ in both populations. Future studies should assemble a range of large samples from worldwide ancestral populations to establish whether these findings are applicable globally. If so, mechanistic studies could narrow in on fewer pathways toward extracting the pathogenesis of SCZ with greater precision.

Methods

EAS sample collection, sequencing and EUR public data collection

We collected 151 prefrontal cortical samples of Han Chinese descent from the National Human Brain Bank for Development and Function according to the standardized operational protocol of China Human Brain Banking Consortium[78,79], and under the approval by the Institutional Review Board of the Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, Beijing, China (Approval Number: 009-2014). We sequenced 151 samples following the BGISEQ-500 protocol outsourced to BGI, WGS and transcriptome sequencing was performed on BGISEQ-500 platform with an average depth of 10X (Supplementary Table 1 and Supplementary Note). To assess differences in ancestry, we also downloaded and processed raw whole genome and RNA-seq data for 407 European ancestry from PsychENCODE/BrainGVEX (Synapse number: syn4590909).

EAS data quality control

Raw sequencing reads were filtered to get clean reads by using SOAPnuke (v1.5.6)[80], and used FastQC to evaluate the quality of sequencing data via several measures, including sequence quality per base, sequence duplication levels, and quality score distribution for each sample. The average quality score for overall DNA and RNA sequences show high scores above 30, indicating that a high percentage of the sequences had high quality (Supplementary Note).

Variant identification

Clean DNA sequencing reads were mapped to the human reference genome hg19 (GRCh37) using BWA-MEM algorithm (BWA v. 0.7.128)[81]. Ambiguously mapped reads (MAPQ <10) and duplicated reads were removed using SAMtools v. 1.29[82] and PicardTools v. 1.130 (http://picard.sourceforge.net/) respectively. Genomic variants were called following the Genome Analysis Toolkit software (GATK v. 3.4–46) best practices[83]. The ancestry of each sample was inferred using data from the 1000 Genomes Project, and no sample was excluded. For EAS cohort, genotypes were imputed into the 1000 Genomes Project phase 3 EAS reference panel by chromosome

using Michigan Imputation Server[84] and subsequently merged. Imputed genotypes were filtered for $R^2 < 0.3$, Hardy-Weinberg equilibrium p-value $< 1 \times 10^{-6}$ and MAF < 0.05, resulting in ~6.4 million autosomal SNPs. For EUR cohort, genotypes were imputed into the HRC reference panel, and removed SNPs with $R^2 < 0.3$, HWE p-value $< 1 \times 10^{-6}$ or MAF < 0.01.

Gene-expression quantification and filter

Mapping of RNA-sequencing reads was completed using STAR (2.4.2a)[85] and the quantification of genes and transcripts was with RSEM (1.3.0)[86]. Raw read counts were log-transformed by R package VOOM first[87], filtering those with log2(CPM)<0 in more than 75% of the samples. We removed all transcripts derived from mitochondrial DNA and X and Y chromosomes. Samples with a Z-score (assessing connectivity between samples) lower than -3 were removed. Quantile normalization was then used to equalize distributions across samples.

Intron cluster quantifications

We used Leafcutter to quantify clusters of variably spliced introns[36]. A cluster consists of overlapping introns that share a splice site. The usage of each intron was first quantified using previously aligned FASTQ files from STAR. Overlapping introns were then grouped with the settings of 50 reads per cluster and a maximum intron length of 500kb.

Co-expression network analysis

To place results from individual genes within their systems-level network architecture, we performed WGCNA[42] using human brain RNA-seq data. Individual (covariate-regressed) expression datasets were combined using the 16,266 genes present across all studies. The resulting normalized mega-analysis expression set was used for all downstream network analyses. We also using robust WGCNA (rWGCNA) to

524 reduce the influence of potential outlier samples on the network architecture. Module
525 robustness was ensured by randomly resampling (2/3 of the total) from the initial set
526 of samples 100 times. This was followed by consensus network analysis, a meta-
527 analytic approach to define modules using a consensus quantile threshold of 0.2. The
528 parameter of rWGCNA was consistent with normal WGCNA (Supplementary Note).

529 **eQTL and sQTL mapping**

530 We used PEER[35] to identify hidden confounders and evaluated the correlation between
531 the known factors (such sex and age) with hidden confounders. We then performed
532 cis-eQTL and cis-sQTL mapping using QTLtools[34], adjusting for PEER factors
533 (Supplementary Note), with a defined cis window of one megabase up- and
534 downstream of the gene/intron cluster body. QTLtools was run in nominal pass mode
535 to detect all available QTLs. QTLtools was also run in the permutation pass mode to
536 identify the best nominal associated SNP per phenotype and with a beta approximation
537 to model the permutation outcome. P-values were then multiple testing corrections
538 using the “q-value” package in R. We define FDR q-value < 0.05 as significant QTL.

539 **Functional enrichment**

540 We performed functional enrichment of both eQTLs and sQTLs using GREGOR[39]
541 (Genomic Regulatory Elements and Gwas Overlap algoRithm) to evaluate the
542 enrichment of variants in genome-wide annotations. GREGOR calculated the
543 enrichment value based on the observed and expected overlap within each annotation.
544 We downloaded the 15-state ChromHMM model BED (Browser Extensible Data) files
545 from the Roadmap Epigenetics Project[88]. We also downloaded 78 consensus
546 transcription factor and DNA-protein binding site BED files existing in multiple cells[89]
547 and then filtered to 50 binding proteins that showed cortical brain expression in EAS
548 and EUR populations data. Lastly, we obtained 171 human RBP site BED files from
549 POSTAR2 database, which was developed as the updated version of CLIPdb and

550 POSTAR and provides the largest collection of RBP binding sites and functional
551 annotations[90].

552 **The fraction of shared eQTL/sQTL between EAS and EUR population**

553 Sharing rate was assessed form all significant eQTLs/sQTLs in the discovery dataset
554 by estimating the proportion of true associations (π_1) on the distribution of
555 corresponding p-values of the overlapping eQTLs /sQTLs in the replication dataset[93].
556 To avoid the influence of sample size in pairwise comparison and get the true
557 replication rate, we randomly selected a subset of European samples (n=151),
558 followed the same pipeline to detect QTLs, and calculated the correlation of genetic
559 effect size of shared eQTL/sQTL between EAS and EUR populations, repeating 100
560 times. We calculated π_1 and used the mean value of π_1 to assess reproducibility
561 between EAS and EUR.

562 **Network preservation analysis**

563 To generate population-specific networks, we compared networks between
564 constructed EAS and EUR populations by individual. We then used WGCNA-
565 integrated function (modulePreservation) to calculate module preservation statistics
566 and applied the Z summary score (Z-score) to evaluate whether a module was
567 conserved or not.

568 **F_{ST} analysis**

569 We used the EAS and EUR panels from the 1000 Genomes Project Phase 3 to
570 investigate the Fixation index (F_{ST}). We estimated F_{ST} using vcftools[91] following the
571 Weir and Cockerham approach[92] for each eSNP and sSNP.

572 We defined population-divergent SNPs as those with F_{ST} >= 0.05 and population-
573 shared SNPs as those with F_{ST} < 0.05. To collect the list of ancestry-specific QTLs and

ancestry-shared QTLs, first, we defined EUR robust QTLs as those that were called significant at least 50 times in down-sampling analysis, as well as ancestry-specific QTLs (significant in one population) and ancestry-shared QTLs (significant in these two populations), by comparing the list of EAS QTLs and EUR robust QTLs. Finally, we performed Fisher's exact test between ancestry-specific QTLs and population-divergent SNPs, as well as population-shared SNPs and ancestry-shared QTLs to test the contribution of AF in QTL comparison.

LD pattern comparison

We calculated the LD score for each SNP as the sum of the LD r^2 between the focal SNP and all flanking SNPs within a 1cM window within the corresponding 1000G EAS and EUR genotype data. We then mapped ancestry-specific eSNPs and sSNPs into LD-score files to obtain the LD score for each ancestry-specific eSNPs or sSNPs in each population. We then performed Wilcox testing to evaluate whether the mean value of the LD score was significantly varied between populations.

Gene-expression prediction

We used matched SNPs and genes from the EAS and EUR populations using matching sample sizes (n=145) to build the gene-expression prediction model. We separated each population into training and validation datasets (100 for training and 45 for validation). Prediction models were built using PrediXcan[15] (Elastic Net) for both populations. Predictive performance (R^2) was measured within each population using nested cross-validation. Wilcoxon tests measured any significant difference in prediction performance across populations.

Partitioned LDSC

Partitioned LD score regression v1.0.1 was used to measure the enrichment of GWAS summary statistics in each functional category by accounting for LD[45]. Brain QTL

599 annotations were created by eSNP and sSNP, mapped to the corresponding 1000
600 Genome reference panel. LD scores were calculated for each SNPs in the QTL
601 annotation using an LD window of 1cM in 1000 Genomes European Phase 3 and 1000
602 Genomes Asian Phase 3 separately. Enrichment for each annotation was calculated
603 by the proportion of heritability explained by each annotation divided by the proportion
604 of SNPs in the genome falling in that annotation category. We then applied Welch
605 Modified Two-Sample t-Test on enrichment values generated from QTLs in the two
606 populations.

607 **Colocalization**

608 We used the conditional association as described in Nica *et al.*[46] to test for evidence of
609 colocalization. This method compares the p-value of association for the lead SNP of
610 an eQTL or sQTL before and after conditioning on the GWAS hits. The equation for
611 the regulatory trait concordance (RTC) Score is as follows: RTC= (N_{SNPs} in an LD
612 block/$Rank_{GWAS\_SNP}$)/ N_{SNPs} in an LD block. The rank denoted the number of SNPs,
613 which when used to correct the expression data, have a higher impact on the QTL than
614 the GWAS SNPs (i.e., $Rank_{GWAS\_SNP}=0$ if the GWAS SNP is the same as the eQTL or
615 sQTL SNP and $Rank_{GWAS\_SNP}=1$ if, of all the SNPs in the interval, the GWAS SNP has
616 the largest impact on the eQTL or sQTL). RTC values close to 1.0 indicated causal
617 regulatory effects. A threshold of 0.9 was used to select causal regulatory elements.

618 **Prioritizing genes underlying GWAS hits**

619 We applied an SMR[9] method on EAS and EUR SCZ GWAS summary data to prioritize
620 candidate genes. We used nominally significant QTLs identified in the previous
621 analysis (FDR < 0.05), containing thousands of unique probes with filtered GWAS
622 summary data ($p < 0.01$) to perform the SMR test. In general, we use the default
623 parameters suggested by the developers of the SMR software. These included the
624 application of heterogeneity independent instruments (HEIDI) testing, filtering out hits

that arose from significant linkage with pleiotropically associated variants (LD cutoff of P = 0.05 in the HEIDI test, as suggested by SMR). Genes with an empirical P passed Bonferroni correction in the SMR test and a P > 0.05 in the HEIDI test were considered as risk genes.

Schizophrenia-related signals

The schizophrenia risk gene sets were collected from publications and databases. For gene analysis, we collected these genes and converted them to Ensembl Gene IDs in Gencode (hg19) using BioMart. We examined whether the risk genes meet one of these criteria: (1) affected by copy number variants (CNVs)[71]; (2) identified by linkage and association study[72-74]; (3) had de novo variants from NP de novo database[75]; (4) identified by convergent functional genomics (CFG)[76]; (5) identified by Pascal gene-based test[76]; or (6) expressed differentially in SCZ[66,77].

Module enrichment

Module functional enrichment of Gene Ontology pathways was assessed with GO-Elite v1.2.5[93] as well as using the clusterprofiler[94] R package, using GO and KEGG databases. For gProfiler, “moderate” hierarchical filtering was used. A custom background set consisted of 10,387 genes present across all studies and microarray platforms. The top pathways were those reaching significance with FDR-adjusted P < 0.05. Module eQTL and candidate genes enrichment were assessed with Fisher exact testing in R.

Acknowledgments

We thank all donors and their families. We thank professor Hailiang Huang for sharing the EAS SCZ GWAS summary statistics. We thank Dr. Elliot Gershon for helpful comments. Tissue was provided by the Human Brain Bank, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China, the Chinese Brain Bank Center, and the Xiangya School of Medicine Brain Bank. This study was supported by National Natural Science Foundation of China (Grants Nos. 31571312, 31970572, 31871276, 91632116 and 81401114), the National Key R&D Project of China (Grants No. 2016YFC1306000 and 2017YFC0908701), Innovation-driven Project of Central South University (Grant Nos. 2020CX003, 2015CXS034 and 2018CX033), Hunan Provincial Natural Science Foundation of China (Grant No. 2019JJ40404), the Fundamental Research Funds for the Central Universities of Central South University, and NIH grants 1 U01 MH103340-01, 1R01ES024988.

Author contributions

S.L. drafted the manuscript, performed the genotype and RNA-seq quality control, mapped eQTL, performed functional enrichment, compared the brain regulatory architecture, as well as integrated QTLs with GWAS signals. Y.C. wrote the manuscript, constructed co-regulatory networks, performed preservation test as well as pathway enrichment analysis. F.W. wrote the manuscript, performed sQTL mapping and functional enrichment analysis. Y. J. preprocessing the genotype and RNA-seq data from PsychENCODE/BrainGVEX project. F.D. and M.L. extracted DNA and RNA, as well as collected sample information. Y.X., R.K, and L.K. substantively revised the manuscript. Z.N. and S.X. participated in the design of comparing the brain regulatory architecture. Sample provided by W.Q., C.M., X.Y., A.B., J.D., J.H. and B.T. C.L. provided PsychENCODE/BrainGVEX data. C.C. conceived, designed and supervised the study and modified the manuscript.

Competing interests

All the authors declare no competing financial interests.

References

1. Hindorff, L.A. *et al.* Prioritizing diversity in human genomics research. *Nat Rev Genet* **19**, 175-185 (2018).
2. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005-D1012 (2019).
3. Rosenberg, N.A. *et al.* Genome-wide association studies in diverse populations. *Nat Rev Genet* **11**, 356-66 (2010).
4. Peterson, R.E. *et al.* Genome-wide association studies in ancestrally diverse populations: Opportunities, methods, pitfalls, and recommendations. *Cell* (2019).
5. Martin, A.R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am J Hum Genet* **100**, 635-649 (2017).
6. Carlson, C.S. *et al.* Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. *PLoS Biol* **11**, e1001661 (2013).
7. Nicolae, D.L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet* **6**, e1000888 (2010).
8. Li, Y.I. *et al.* RNA splicing is a primary link between genetic variation and disease. *Science* **352**, 600-4 (2016).
9. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48**, 481-7 (2016).
10. Hormozdiari, F. *et al.* Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am J Hum Genet* **99**, 1245-1260 (2016).
11. Joehanes, R. *et al.* Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol* **18**, 16 (2017).
12. Marigorta, U.M. *et al.* Transcriptional risk scores link GWAS to eQTLs and predict complications in Crohn's disease. *Nat Genet* **49**, 1517-1521 (2017).
13. Brown, A.A. *et al.* Predicting causal variants affecting expression by using whole-genome sequencing and RNA-seq from multiple human tissues. *Nat Genet* **49**, 1747-1751 (2017).
14. Hormozdiari, F. *et al.* Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. *Nat Genet* **50**, 1041-1047 (2018).
15. Gamazon, E.R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* **47**, 1091-8 (2015).
16. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245-52 (2016).
17. BrainSeq: Neurogenomics to Drive Novel Target Discovery for Neuropsychiatric Disorders. *Neuron* **88**, 1078-1083 (2015).
18. De Jager, P.L. *et al.* Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat Neurosci* **17**, 1156-63 (2014).
19. Psych, E.C. *et al.* The PsychENCODE project. *Nat Neurosci* **18**, 1707-12 (2015).
20. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
21. Mikhaylova, A.V. & Thornton, T.A. Accuracy of Gene Expression Prediction From Genotype Data With PrediXcan Varies Across and Within Continental Populations. *Front Genet* **10**,

261 (2019).
22. Mogil, L.S. *et al.* Genetic architecture of gene expression traits across diverse populations. *PLoS Genet* **14**, e1007586 (2018).
23. Gottlieb, A. *et al.* Cohort-specific imputation of gene expression improves prediction of warfarin dose for African Americans. *Genome Med* **9**, 98 (2017).
24. Stranger, B.E. *et al.* Patterns of cis regulatory variation in diverse human populations. *PLoS Genet* **8**, e1002639 (2012).
25. Mele, M. *et al.* Human genomics. The human transcriptome across tissues and individuals. *Science* **348**, 660-5 (2015).
26. Gamazon, E.R. *et al.* Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat Genet* **50**, 956-967 (2018).
27. Jiang, L. *et al.* DESE: estimating driver tissues by selective expression of genes associated with complex diseases or traits. *Genome Biol* **20**, 233 (2019).
28. Finucane, H.K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet* **50**, 621-629 (2018).
29. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).
30. Xiang, R. *et al.* Genome variants associated with RNA splicing variations in bovine are extensively shared between tissues. *BMC Genomics* **19**, 521 (2018).
31. Collado-Torres, L. *et al.* Regional Heterogeneity in Gene Expression, Regulation, and Coherence in the Frontal Cortex and Hippocampus across Development and Schizophrenia. *Neuron* **103**, 203-216 e8 (2019).
32. Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci* **19**, 1442-1453 (2016).
33. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**(2018).
34. Delaneau, O. *et al.* A complete tool set for molecular QTL discovery and analysis. *Nat Commun* **8**, 15452 (2017).
35. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc* **7**, 500-7 (2012).
36. Li, Y.I. *et al.* Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet* **50**, 151-158 (2018).
37. Kim, Y. *et al.* A meta-analysis of gene expression quantitative trait loci in brain. *Transl Psychiatry* **4**, e459 (2014).
38. Walker, R.L. *et al.* Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. *Cell* **179**, 750-771 e22 (2019).
39. Schmidt, E.M. *et al.* GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics* **31**, 2601-6 (2015).
40. Consortium, G.T. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213 (2017).
41. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
42. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
43. Lam, M. *et al.* Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat Genet* **51**, 1670-1678 (2019).
44. Bipolar, D., Schizophrenia Working Group of the Psychiatric Genomics Consortium. Electronic address, d.r.v.e., Bipolar, D. & Schizophrenia Working Group of the Psychiatric Genomics, C. Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28

Subphenotypes. *Cell* **173**, 1705-1715 e16 (2018).
45. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
46. Nica, A.C. *et al.* Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet* **6**, e1000895 (2010).
47. Takata, A., Matsumoto, N. & Kato, T. Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. *Nat Commun* **8**, 14519 (2017).
48. Raj, T. *et al.* Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. *Nat Genet* **50**, 1584-1592 (2018).
49. Narahara, M. *et al.* Large-scale East-Asian eQTL mapping reveals novel candidate genes for LD mapping and the genomic landscape of transcriptional effects of sequence variants. *PLoS One* **9**, e100924 (2014).
50. Wang, X. *et al.* Mapping of hepatic expression quantitative trait loci (eQTLs) in a Han Chinese population. *J Med Genet* **51**, 319-26 (2014).
51. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506-11 (2013).
52. Fraser, H.B., Lam, L.L., Neumann, S.M. & Kobor, M.S. Population-specificity of human DNA methylation. *Genome Biol* **13**, R8 (2012).
53. Tehranchi, A. *et al.* Fine-mapping cis-regulatory variants in diverse human populations. *Elife* **8**(2019).
54. Cann, H.M. *et al.* A human genome diversity cell line panel. *Science* **296**, 261-2 (2002).
55. International HapMap, C. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52-8 (2010).
56. Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
57. Novembre, J. & Di Rienzo, A. Spatial patterns of variation due to natural selection in humans. *Nat Rev Genet* **10**, 745-55 (2009).
58. Barreiro, L.B., Laval, G., Quach, H., Patin, E. & Quintana-Murci, L. Natural selection has driven population differentiation in modern humans. *Nat Genet* **40**, 340-5 (2008).
59. Torres, J.M. *et al.* Cross-tissue and tissue-specific eQTLs: partitioning the heritability of a complex trait. *Am J Hum Genet* **95**, 521-34 (2014).
60. Disease, G.B.D., Injury, I. & Prevalence, C. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **392**, 1789-1858 (2018).
61. Cai, L. *et al.* Implications of Newly Identified Brain eQTL Genes and Their Interactors in Schizophrenia. *Mol Ther Nucleic Acids* **12**, 433-442 (2018).
62. Romme, I.A., de Reus, M.A., Ophoff, R.A., Kahn, R.S. & van den Heuvel, M.P. Connectome Disconnectivity and Cortical Gene Expression in Patients With Schizophrenia. *Biol Psychiatry* **81**, 495-502 (2017).
63. Sekar, A. *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature* **530**, 177-83 (2016).
64. Schadt, E.E. Molecular networks as sensors and drivers of common human diseases. *Nature* **461**, 218-23 (2009).
65. Liu, Y. & Chance, M.R. Pathway analyses and understanding disease associations. *Curr Genet Med Rep* **1**(2013).
66. Gandal, M.J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**(2018).
67. Kushima, I. *et al.* High-resolution copy number variation analysis of schizophrenia in Japan. *Mol Psychiatry* **22**, 430-440 (2017).

68. Kushima, I. *et al.* Comparative Analyses of Copy-Number Variation in Autism Spectrum Disorder and Schizophrenia Reveal Etiological Overlap and Biological Insights. *Cell Rep* **24**, 2838-2856 (2018).
69. Berdenis van Berlekom, A. *et al.* Synapse Pathology in Schizophrenia: A Meta-analysis of Postsynaptic Elements in Postmortem Brain Studies. *Schizophr Bull* **46**, 374-386 (2020).
70. Hashimoto, R. *et al.* A missense polymorphism (H204R) of a Rho GTPase-activating protein, the chimerin 2 gene, is associated with schizophrenia in men. *Schizophr Res* **73**, 383-5 (2005).
71. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237-41 (2008).
72. Ng, M.Y. *et al.* Meta-analysis of 32 genome-wide linkage studies of schizophrenia. *Mol Psychiatry* **14**, 774-85 (2009).
73. Lewis, C.M. *et al.* Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet* **73**, 34-48 (2003).
74. Allen, N.C. *et al.* Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet* **40**, 827-34 (2008).
75. Li, J. *et al.* Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database. *Mol Psychiatry* **21**, 298 (2016).
76. He, X. *et al.* Sherlock: detecting gene-disease associations by matching patterns of expression QTL and GWAS. *Am J Hum Genet* **92**, 667-80 (2013).
77. Gandal, M.J. *et al.* Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693-697 (2018).
78. Yan, X.X., Ma, C., Bao, A.M., Wang, X.M. & Gai, W.P. Brain banking as a cornerstone of neuroscience in China. *Lancet Neurol* **14**, 136 (2015).
79. Qiu, W. *et al.* Standardized Operational Protocol for Human Brain Banking in China. *Neurosci Bull* **35**, 270-276 (2019).
80. Chen, Y. *et al.* SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* **7**, 1-6 (2018).
81. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
82. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
83. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
84. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287 (2016).
85. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
86. Li, B. & Dewey, C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
87. Ritchie, M.E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
88. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
89. Arbiza, L. *et al.* Genome-wide inference of natural selection on human transcription factor binding sites. *Nat Genet* **45**, 723-9 (2013).
90. Zhu, Y. *et al.* POSTAR2: deciphering the post-transcriptional regulatory logics. *Nucleic Acids Res* **47**, D203-D211 (2019).
91. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-8 (2011).
92. Balloux, F. & Lugon-Moulin, N. The estimation of population differentiation with microsatellite markers. *Mol Ecol* **11**, 155-65 (2002).

869 93. Zambon, A.C. *et al.* GO-Elite: a flexible solution for pathway and ontology over-
870 representation. *Bioinformatics* **28**, 2209-10 (2012).
871 94. Yu, G., Wang, L.G., Han, Y. & He, Q.Y. clusterProfiler: an R package for comparing
872 biological themes among gene clusters. *OMICS* **16**, 284-7 (2012).

873 Data availability

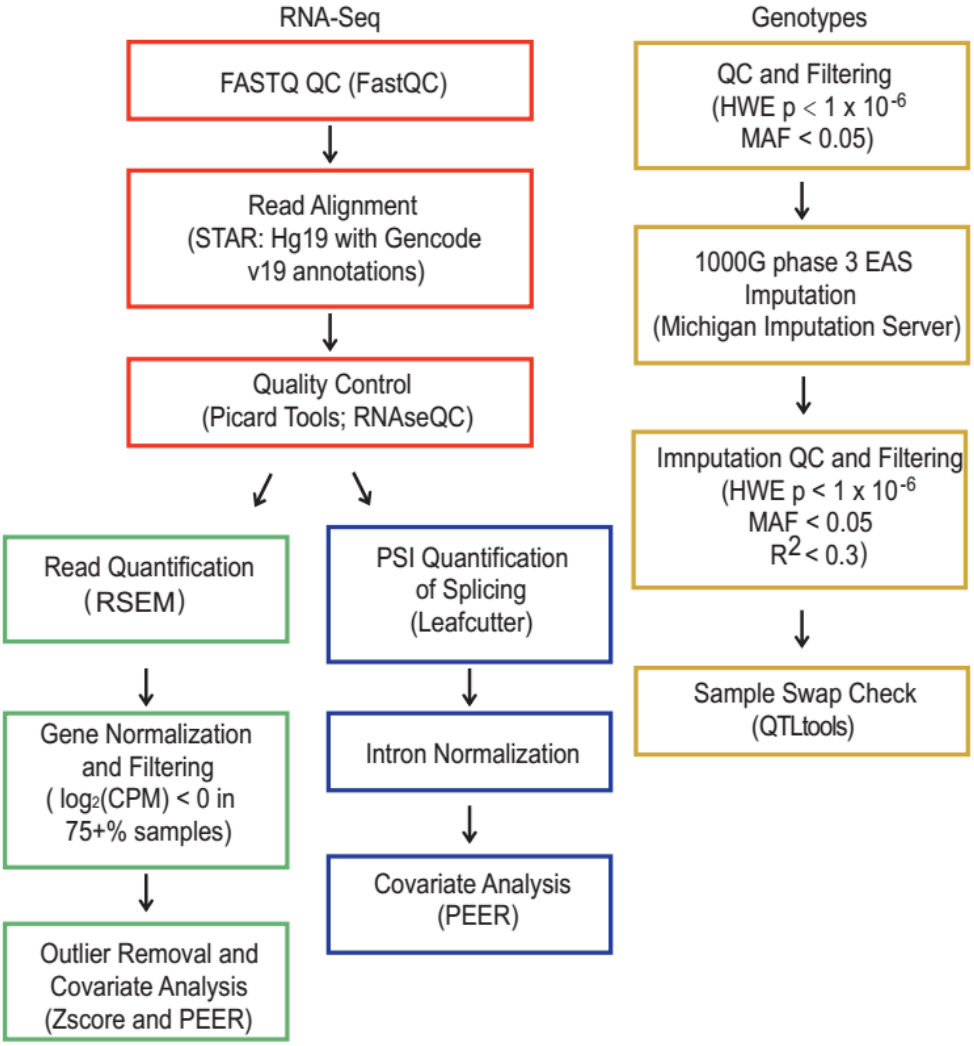
874 The raw sequence data reported in this paper have been deposited in the Genome
875 Sequence Archive in BIG Data Center, Beijing Institute of Genomics (BIG), Chinese
876 Academy of Sciences, under accession numbers HRA000108, HRA000108 that can
877 be accessed at https://bigd.big.ac.cn/gsa-human. eQTL and sQTL summary results
878 for EAS samples can be downloaded from
879 http://brainexpnpd.org:8088/BrainEXPNPD/download.html.

880 Code availability

881 Codes are available at https://github.com/liusihan/population-compare-pipeline

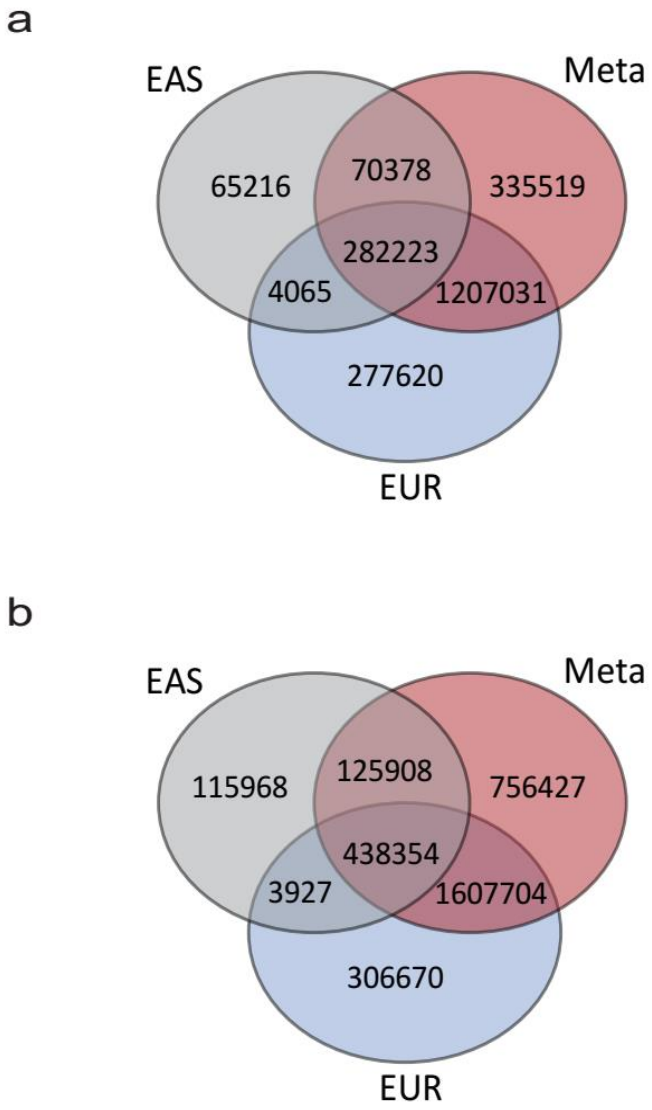
882 Additional information

883 Correspondence and requests for materials should be addressed to C.C.



884

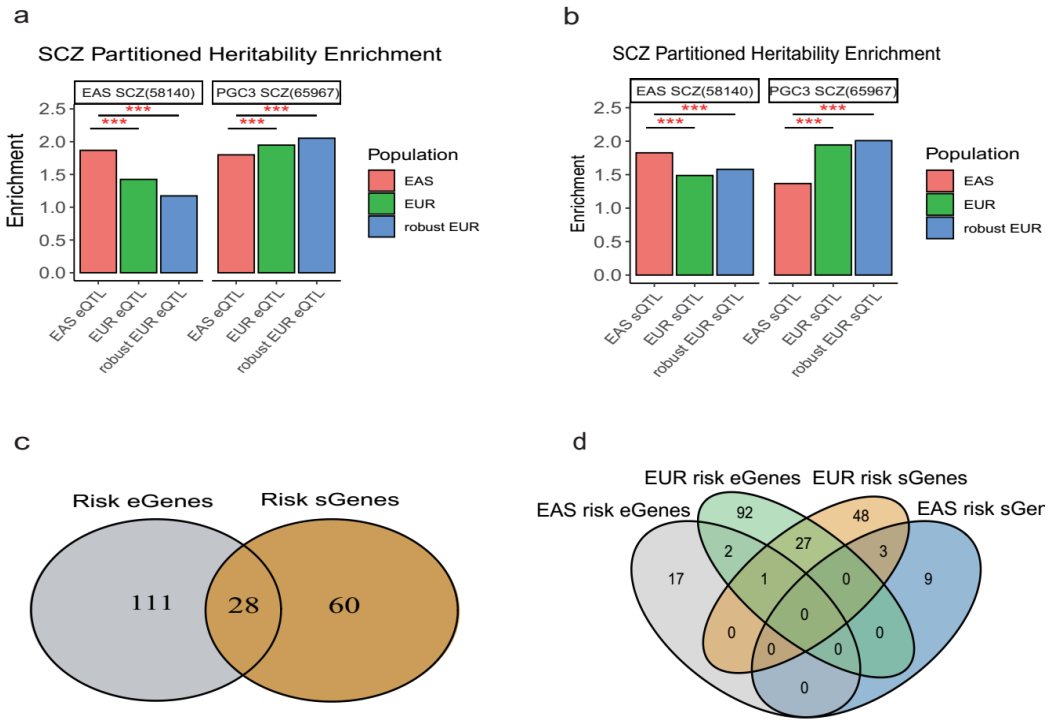
885 **Extended Data Fig.1 | Overview of methods and QC pipeline for EAS samples.**



886

887 **Extended Data Fig.2 | QTL comparison. a,** Venn plot for eQTLs. **b,** Veen plot for

888 sQTLs.



e

GO term	Description	p-value	p.adjust	Count
GO:0098984	neuron-to-neuron synapse	7.04E-11	2.95E-08	36
GO:0014069	postsynaptic density	1.20E-10	2.95E-08	34
GO:0032279	asymmetric synapse	1.74E-10	2.95E-08	34
GO:0099572	postsynaptic specialization	3.08E-10	3.91E-08	35
GO:0051056	regulation of small GTPase mediated signal transduction	5.18E-09	1.40E-05	33
GO:0050804	modulation of chemical synaptic transmission	1.03E-08	1.40E-05	37
GO:0099177	regulation of trans-synaptic signaling	1.10E-08	1.40E-05	37
GO:0051020	GTPase binding	2.59E-08	1.66E-05	39
GO:0060589	nucleoside-triphosphatase regulator activity	1.58E-07	3.38E-05	28
GO:0031267	small GTPase binding	1.58E-07	3.38E-05	32

889

890 **Extended Data Fig.3 | Integrating SCZ GWAS signals**. **a,** GWAS signals enrichment
891 comparison for eQTLs. **b,** GWAS signals enrichment comparison for sQTLs. **c-d,** Venn
892 plot for risk genes identified by eQTLs and sQTLs. **e**, List of the top ten pathways which
893 enriched in SCZ risk genes. p.adjust: Bonferroni adjusted p-value. Count: number of
894 genes located in this pathway.